Region One
Vegetation Classification, Mapping,
Inventory and Analysis Report

Numbered Report 11-02 **April, 2011**

The Region 1 Existing Vegetation Mapping Program (VMap) Beaverhead-Deerlodge Methodology

Stephen R. Brown Jr.
Robert S. Ahl

USDA Forest Service, Region 1
Engineering Staff
200 E. Broadway, Missoula, MT 59802

1. Data

Image Data

The two cost effective data sets available for use were Landsat Thematic Mapper (TM) and National Agriculture Imagery Program (NAIP) imagery. Both of these datasets have properties that add value to the mapping work. Although the spatial resolution of Landsat TM is coarse at 30 meters, it has spectral properties that exceed what are offered in the high resolution NAIP imagery. These spectral properties are useful in discerning different types of vegetation and vegetative state. Landsat has a quick return cycle, repeating inventory of the same area every 16 days. The footprint of Landsat is large (each scene covering approximately 185 sq kilometers) which makes for consistent data sets over large areas — important for larger area projects such as the B-D VMap. Because of the high spatial resolution (1 meter), the NAIP imagery available offered properties that the Landsat could not. The polygon or map units delineated from these data (see section on 'image segmentation' for a further description of this process) are very accurate as compared to what can be accomplished from Landsat, and the secondary statistics derived from this imagery are useful for better delineation of the various cover types that are mapped. The downside of this imagery are that each image is small (approximately 20,000 acres) and so digital numbers can vary across larger areas. Models (see section on 'Model Areas' for description of models) were kept small in the project to minimize this source of error. The following are the imagery used in the B-D VMap project.

- Landsat Thematic Mapper imagery: A mid-summer image (July/August) was selected to capture "peak green" vegetation fully mature prior to senescence. The Landsat TM images (TM bands 1,2,3,4,5, and 7) used in the project for each model were collected in August of 2009. All TM images were orthorectified to the color infrared NAIP imagery used in the project and radiance reflectance corrected.

- 1 meter National Agriculture Imagery Program (NAIP) color infrared and natural imagery: NAIP imagery used in the project are color infrared (IR) digital orthorectified photos of Montana with the images acquired in summer 2009. The original digital images were then processed to an IR product with a 1 meter ground sample distance (GSD) and rectified to National Mapping Standards at the 1:24,000 scale. This imagery was sampled back to 5m using ERDAS Imagine and used for segmentation and image processing where appropriate. Visit http://gisportal.mt.gov/Portal/DiscoveryServlet for complete metadata on this imagery.

Ancillary Data and Image Derivatives

In addition to the imagery, the following were also important inputs into the mapping process. All of these were produced from the Landsat or NAIP imagery or 10 meter Digital Elevation Models (DEM's).

1. **10 meter National Elevation Data (NED)—** (received from the Forest Service Remote Sensing Applications Center) were used in eCognition for species modeling and also used to classify portions of the non-forest cover (see 'methodology' section for classes

mapped) with the Random Forest tree predictors using Rv2.7.2 (see ‘methodology section for useage of Random Forest’s in VMap project.

2. **Solar Radiation** — used to model non-forest cover (see methodology for the cover types mapped) and to create the individual tree dominance type probability surfaces. The function is part of ArcGIS Spatial Analyst and is a calculation of how much sun an area receives over a period of time. Since radiation can be greatly affected by topogra phy and surface features, a key component of the calculation algorithm requires genera tion of an upward looking hemispherical viewshed for every location in a digital eleva tion model. The input DEM used or this calculation was 30 meter National Elevation Data (did not use 10 meter for computation restrictions) using all default parameters. Time configuration was set to “whole year with monthly intervals” for Year 2005. See http://webhelp.esri.com/arcgisdesktop/9.2/index.cfmTopicName=Area_Solar_Radiation for a more detailed description of this function.

3. **TRASP**—- used to model non-forest cover and to create the individual tree dominance type probability surfaces. The circular aspect variable is transformed to a radiation index (TRASP) in this calculation. This transformation assigns a value of zero to land ori ented in a north-northeast direction, (typically the coolest and wettest orientation), and a value of one on the hotter, dryer south-southwesterly slopes.The result is a continuous variable between 0 - 1 (Roberts and Cooper 1989).

$$TRASP= \frac{1 - \cos((pi / 180)(aspect - 30))}{2}$$

4. **Slope—-** used to model non-forest cover and to create the individual tree dominance type probability surfaces. Percent Slope calculated from the 10 meter NED data using ER DAS Imagine.

5. **CTI**—-- used to model non-forest and to create the individual tree dominance type prob ability surfaces.CTI is a steady state wetness index. The CTI is a function of both the slope and the upstream contributing area per unit width orthigonal to the flow direction. CTI was desigined for hillslope catenas. Accumulation numbers in flat areas will be very large and CTI will not be a relevant variable. CTI is highly correlated with several soil attributes such as horizon depth(r=0.55), silt percentage(r=0.61), organic matter content(r=0.57), and phosphorus(r=0.53) (Moore et al. 1993).

The implementation of CTI can be shown as:

$$CTI = \ln (As / (\tan (beta))$$

where As = Area Value calculated as (flow accumulation + 1) * (pixel area in m2) and beta is the slope expressed in radians.

6. **Minimum Texture—-** used in eCognition for segmentation of each model and for model ing. The texture image characterizes the spatial homogeneity or heterogeneity of each

pixel based on its surrounding neighbors. Minimum texture, developed by Woodcock and Ryherd (1996), calculates a minimum variance from an adaptive window around each pixels as it's measure of the texture. The resulting texture image is a composite of the minimum variance values calculated for each pixel. As shown by Coburn and Rob erts (2004), three bands of image texture can improve classification overall by 13 percent with 4 to 8 percent improvement when compared to use of a single band of texture. The image used to create texture for each of the models was a principal component image of the 1m color infrared NAIP imagery. The texture bands were created using adaptive windows of 5x5, 15X15, and 25X25 and the resultant texture images were then resampled to 5m.

7. **Mean Texture**—- used in eCognition for segmentation and classification of the models. Mean texture was calculated from a 5m principal component of each CIR NAIP image using a mean variance using adaptive windows of 3X3, 5X5, and 9X9.

8. **Tassel Cap Transformations**—-Tassel Cap (TC) was calculated for all the Landsat imagery used in eCognition modeling process . TC is a linear transformation of the reflectance calculated TM data that rotates the data structure such that the majority of the information contained in the 6 bands will occupy 3 dimensions that are directly related to the on-the-ground physical scene characteristics (Kauth and Thomas, 1976.). These dimensions define planes of soils (brightness), vegetation (greenness), and a transitional zone that relates to canopy and soil moisture (wetness). These three dimensions capture 97%+ of the data variation in the 6 TM bands and can enable the discernment of key forest attributes (i.e., species, age, and structure.)

9. **NDVI—**The Normalized Difference Vegetation Index (NDVI) is calcu lated as the normalized difference between the NIR and the Red bands (NIR - R)/(NIR + R). NDVI was used in the eCognition modeling process. The NDVI is probably the most widely used vegetation index and has been shown to be related to a number of different biomass variables. Simple vegetation indices such as NDVI, however, provide an inadequate representation of complex vegetation cover as they are related only to the total amount of above-ground green leaf biomass, and give no indication of the types of vegetation present. Vegetated areas will generally yield a higher NDVI value than rock, which will have values greater than that of clouds, snow, and water. The 5meter NAIP imagery was used to calculate NDVI and used in eCognition where applicable. In other cases, NDVI was calculated for the Landsat TM imagery was used in the modeling process.

9. **Principal Component—**used in eCognition modeling. This was calculated using ERDAS Imagine's function to create six bands of principal component. Principal component (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal component. The first principle component accounts for as much of the variability in the data as possible. PCA has proven to be of value in analysis of multispectral data (Press et al. 1992.) The transformation of the raw remote sensor data using PCA can result in images more interpretable than the original data.

2. Modeling Unit Construction

Model Areas

To make the 30 meter Landsat TM and the 1 meter NAIP data useable for image processing, both sets of data were resampled to 5 meters using a cubic convolution resampling. At 5m resolution, data sets are still quite large, and given the limits of the image processing software used (eCognition), model areas were created no larger than 450,000 acres. The larger basis for the stratification of these models coincides with the USDA Forest Service National Hierarchical Framework of Ecological Units (Bailey et al. 1994.) Beyond these units, models were further subdivided to Forest Service district or other management units and kept to a size manageable by the software used to process it. Fifteen models were created to cover the entire B-D National Forest, though two models (2901 and 2902) were mapped previously as part of the Eastside VMap.

Image Segmentation

Image segmentation is the process of combining the pixels within digital images into spatially cohesive units, or regions, thereby creating image objects. These objects represent discrete areas in the image. This segmentation and merging process is influenced by the variance structure of the image data and provides the modeling units that reflect life form composition, stocking, tree crown size differences, and other vegetation and/or landcover characteristics (Haralick and Shapiro 1985, Ryerd and Woodcock 1996). Image segmentation was used in the eastside VMap to delineate vegetative features using Definiens' eCognition software version 4.06. The segmentation process in eCognition is based on both the local variance structure within imagery and shape indices. These image objects effectively depict the elements of vegetation and landcover pattern on the landscape (McDonald et al. 2002.) Figure (3) illustrates the image segmentation-based depiction of landscape pattern displayed over 1m NAIP imagery.

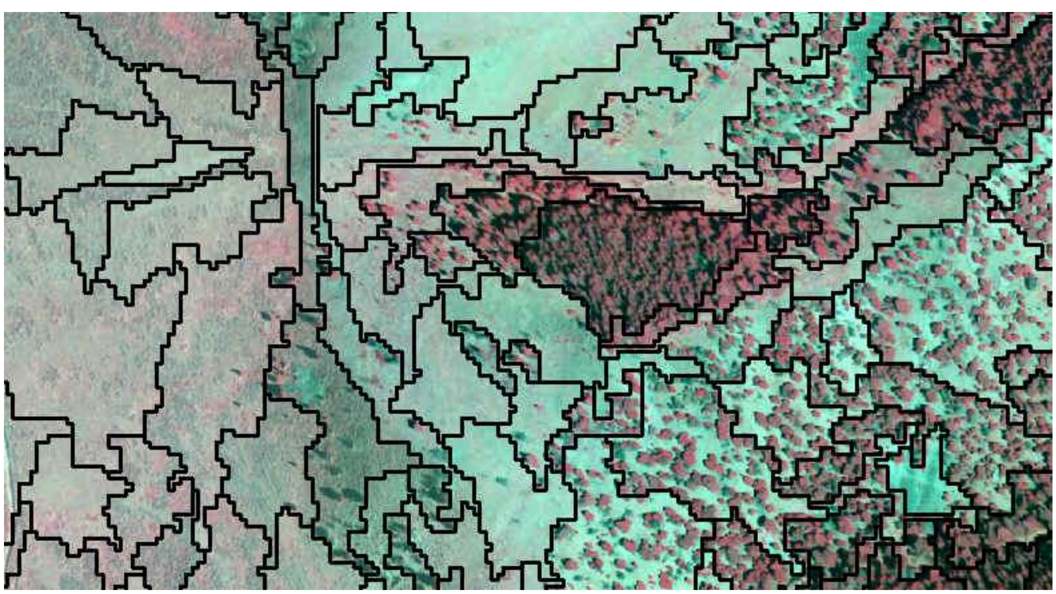


Figure 3. Example of image segmentation from eCognition Software (image segmentation draped over 1m color infrared NAIP imagery.)

3. Training Data

Any remote sensing product is only as good as the ground data associated with it. Training data is used to build the relationships between ground based phenomena and the spectral information contained in a remotely sensed image. Using these known areas, then, it is possible to construct algorithms to predict and label the unknown areas within a study area.

For the project, the image objects, or polygons, were the units used to collect each training site. Collection of the training data was completed using various methods but primarily it was a ground-based sampling method. In previous mapping efforts training data collection relied heavily on the use of aerial photography. On the B-D, however, the photos are almost 20 years old and were not deemed current enough to provide reliable training data in light of the recent beetle infestations. Some tree data could be interpreted from the 1m NAIP if personnel were familiar with the area (i.e., Dominance Type and Tree Canopy Cover). All of the non-forest data was collected from field sampling (contact Northern Region Geospatial Group for non-forest field sampling techniques.)

LANDSCAPE STRATIFICATION

One of the primary goals of our field activities is to capture the variation of landscape characteristics occurring on the Beaverhead-Deerlodge National Forest (B-D NF). To begin our interpretation of landscape variability we gathered individual datasets representing the climatic, geologic, vegetative, and topographic characteristics encompassing the B-D National Forest (Table 1). The data were assembled from a variety of sources, and their original format spanned multiple scales and data models.

Table 1. *Landscape stratification input data, sources, and resolution*

Characteristic	Source	Type	Resolution
Topography	National Elevation Dataset	Raster	30m
Precipitation	DAYMET	Raster	1000m
Heat Units	DAYMET	Raster	1000m
Land Type Association	R1 LTA	Vector	???
State of MT Geology	MT NRIS	Vector	1:500,000
Ecoregion, L4	EPA	Vector	
Vegetation	SILC3	Raster	30m

For use in our landscape stratification scheme, the data were all converted to raster format, with 30 meter grid cells. We found this resolution to be appropriate because it afforded sufficient detail over the roughly 5,000,000 acres comprising the B-D, while also being computationally efficient. Vegetation modeling units of the B-D NF are given below in Figure 1.

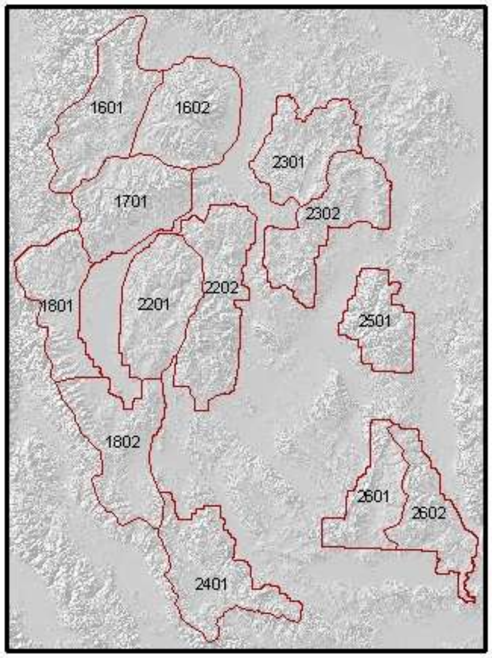


Figure 1. *Vegetation modeling units within the Beaverhead-Deerlodge National Forest*

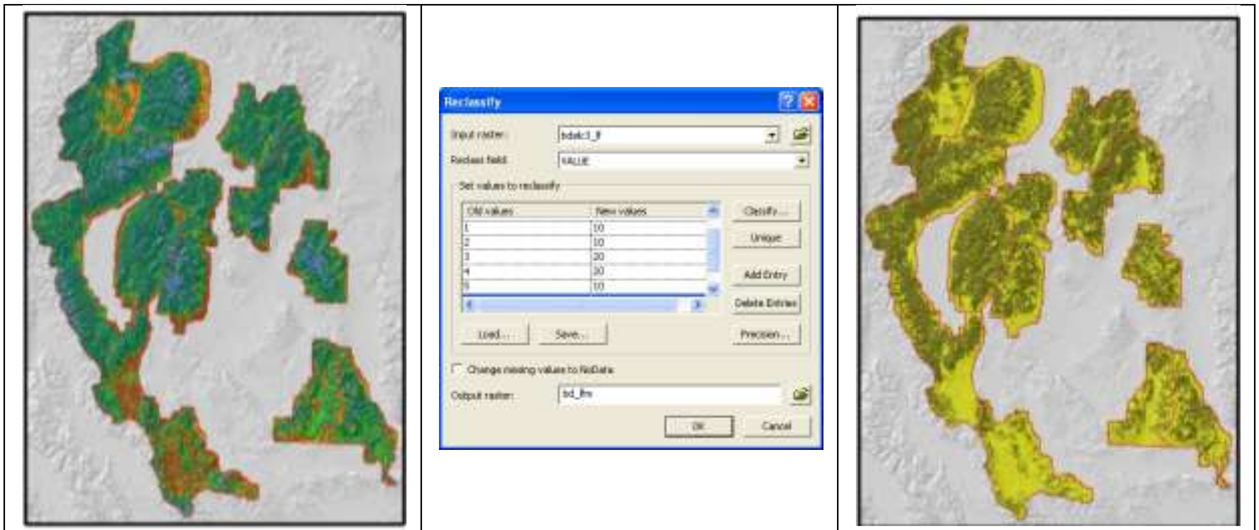
When interpreting the various data we focused on landscape characteristics that have the potential to influence vegetation patterns. For instance, elevation, precipitation, temperature, and parent materials are elemental drivers of vegetative distributions. Many of the layers used to describe biophysical properties of the landscape present data in a continuous fashion. To generalize the provided level of detail, layers of continuous data were reclassified into broad but meaningful ranges. For example, data from the National Elevation Dataset (NED) originally provided continuous elevation estimates rounded to the nearest foot, but this level of detail was difficult to work with. We therefore reclassified the dataset into three classes, essentially representing low, medium, and high elevation landscape units. A similar procedure was also applied to the precipitation and temperature datasets, and is illustrated below using elevation as an example. The Natural Breaks classification algorithm was used to parse the elevation histogram into the specified number of classes, which in the example below was three (Figure 2).

A) Continuous Elevation in ft.	**B**) Classification Algorithm	**C**) Low, Medium, High Elevation

Figure 2. C*lassification of continuous elevation data using the natural breaks algorithm to produce three classes ranging from 1) 3,000 – 6,500 2) 6,501 -8,00 3) and 8,000 – 11, 395 ft, shown in brown, blue, and pink, respectively*

A high degree of correlation between the elevation, precipitation, and temperature datasets was detected. Considering the degree of layer correlation and that elevation was a variable in the algorithm used to create the climate estimates, we chose to use the pure elevation dataset for landscape classification. As such, elevation data allowed us to vertically stratify the landscape into three general classes, representing low, moderate, and high elevation zones.

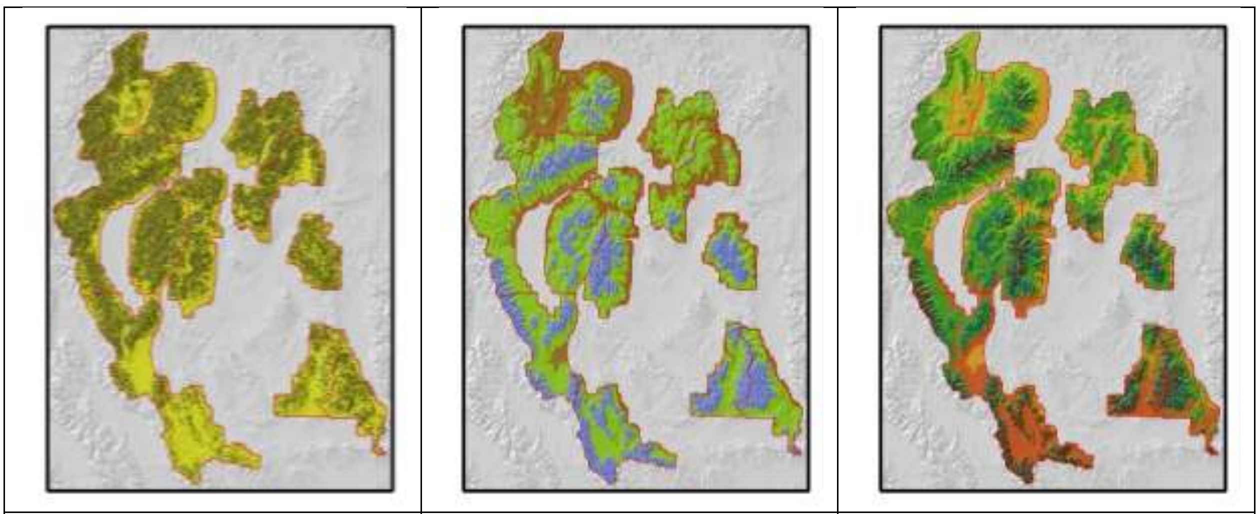
Further division of the landscape focused on the horizontal distribution of features. While we considered mapped distributions of geomorphic land types and their various associations (R1 LTA), regional geology, and Level 4 Ecoregion data layers, we found a basic classification of forest versus non-forest lifeforms to be the most meaningful and straight-forward in its interpretation. As such, the latest Satellite Image Landcover Classification (SILC3) dataset was reduced into two categories describing the basic forest and non-forest lifeforms within the B-D NF.



A) SILC3 with all Veg. Classes	**B**) Classification Algorithm	**C**) SILC3 Reclassified to Lifeform

Figure 3. *Reclassification of the SILC3 herbaceous, dry shrub, deciduous tree, coniferous tree, and rock types into a basic forest versus non-forest lifeform map*.

We arrived at our final land unit stratification by combining both the vertical and horizontal elements of the landscape. The vertical elements represented the low, moderate and high elevation classes, and the horizontal elements were composed of forest and non-forest vegetation types.



A) Lifeform with 2 classes	**B**) Elevation with 3 classes	**C**) ELE-LFM with 6 classes

Figure 4. *Development of the final landscape stratification dataset based on forest and non-forest lifeforms and elevation zones. Two classes of A) lifeform were combined with three classes of B) elevation to create 6 unique combinations (strata) of vertical and horizontal landscape features.*

By sampling vegetation within the unique combinations of forest and non-forest types over a range of elevation classes should ensure that we cover the range of expected environmental conditions in the B-D landscape.

Table 2. *Description and interpretation of the integrated Beaverhead-Deerlodge NF Strata map inputs and final unique combinations of lifeform and elevation classes.*

STRATA	Lifeform Input	Elevation Input	Interpretation
11	Class 10 (Non-Forest)	Class 1 (0 - 6,500 ft)	Non-Forest at low elevation
12	Class 10 (Non-Forest)	Class 2 (6,500 - 8,000 ft)	Non-Forest at mederate elevation
13	Class 10 (Non-Forest)	Class 3 (8,000 - 11,395 ft)	Non-Forest at high elevation
21	Class 20 (Forest)	Class 1 (0 - 6,500 ft)	Forest Lifeform at low elevation
22	Class 20 (Forest)	Class 2 (6,500 - 8,000 ft)	Forest Lifeform at moderate elevation
23	Class 20 (Forest)	Class 3 (8,000 - 11,395 ft)	Forest Lifeform at high elevation

Table 3. *Spatial characteristics of Beaverhead-Deerlodge NF vegetation modeling units and associated Strata groups*.

			Percent of STRATA in Model					
MODEL	**TOTAL ACRES**	**STRATA Groups**	**11**	**12**	**13**	**21**	**22**	**23**
1601	425,745	6	24	2	1	37	35	2
1602	394,211	6	26	3	2	23	32	7
1701	382,113	6	9	3	10	12	51	15
1801	306,390	6	6	7	6	8	66	7
1802	412,133	5	7	32	8	0	29	24
2201	388,696	6	8	9	2	6	54	21
2202	447,796	6	15	14	8	6	37	20
2301	323,910	6	18	3	0	23	54	2
2302	374,763	6	20	9	1	25	42	3
2401	405,589	6	3	47	32	0	8	10
2501	230,892	6	11	10	9	9	36	25
2601	261,136	6	3	34	26	0	19	17
2602	288,802	6	7	30	17	2	27	17
	4,642,177	average proportion	12	16	9	12	38	13

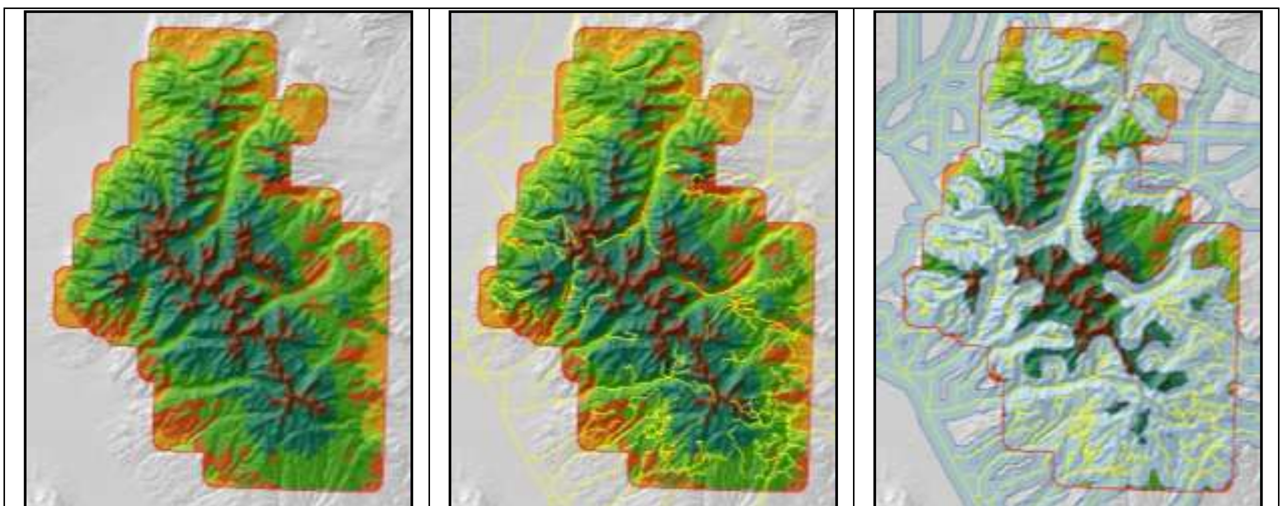
SAMPLING WITHIN STRATA

Upon development of the biophysical strata composing the B-D NF model areas, the next stage of the VMAP sampling strategy is to identify potential sites for field review. There are three essential considerations in the development of a proposed sample network. First, would like to distribute our sample network proportionately across the landscape. Second, it is desirable to collect as many good samples as possible. In keeping with the first two principles, the time and effort needed to access suggested sample sites must be balanced against the need to acquire a certain number of samples. In short, spending excessive effort to visit a few remote sample sites may not be as efficient as collecting more, but easier to obtain samples.

To set up a spatially proportionate sample design, we created a systematic grid of points with 500 meter spacing across the entire study area, where each point represents a potential field review site. Each point was attributed with a vegetation model identification number, and relevant Strata code. The basic assumption is that if all potential sites are reviewed, a proportionate sample of landscape features and associated vegetation characteristics will be sampled. Given that it will not be possible to visit all sites, further stratification is necessary to derive a realistic proposed sample network.

Sample Reduction

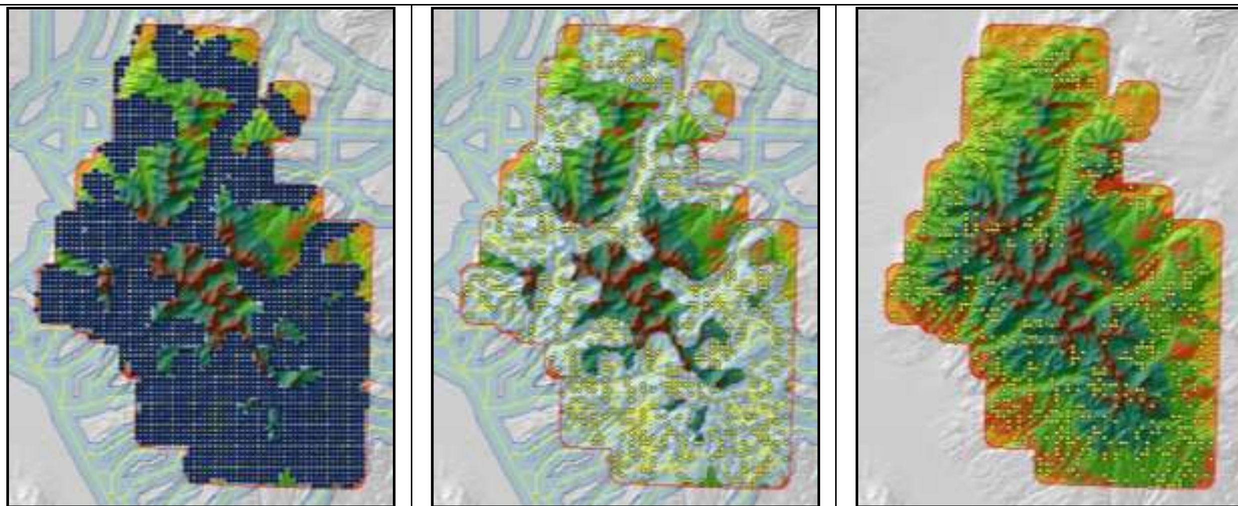
As a first step towards reducing the potential sample points down to a reasonable number we assumed that the existing road network will determine our primary access to proposed sites. Give that the amount of time required to obtain and record sample data is limited, we applied a 1 km buffer (about 0.5 mile) buffer around the road network. The zone identified by the buffer network then represents potential areas within vegetation modeling units that may be visited by a sample collection crew with a reasonable amount of effort. An example of this buffer network is given below for the Tobacco Root Mountains vegetation sub-model (m2501) in Figure 5.



A) Model 2501 STRATA	**B**) STRATA with Roads	**C**) STRATA with Buffered Road

Figure 5. *Sample reduction, phase 1*

While collection of sample data from sites outside of the road buffer zone may be valuable, the amount of time and effort to reach them may be excessive. We therefore focused our proposed sample network on sites within the buffer network. This reduced the number of possible sample points from 75,067 to 44,822. Despite a roughly 50% reduction, 44,822 points still represents approximately 3,500 sample sites for each vegetation modeling unit, and this is still more than our initial sampling effort can accomplish in a field season. To further reduce the potential sample network, we randomly selected 25% of buffered points within each Strata, within each vegetation modeling unit of the B-D NF. This resulted in a network of 11,207 suggested sample points across the entire Beaverhead-Deerlodge National Forest, with a minimum of 584, maximum of 1,282, and mean of 866 locations in each of the 13 vegetation modeling units we intend to review in the 2009 field season. The process of selecting points within the buffer zone is illustrated schematically below in Figure 6, using the Tobacco Root Mountains vegetation sub-model (m2501) as an example.



A) Full Network of Systematic Points within the 1km Buffer	**B**) Random Selection of 25% of Buffered Points in Each Strata	**C**) Final Illustration of Suggested Sample Sites in Model 2501

Figure 6. *Sample reduction, phase 2*.

Within Tobacco Root Mountains vegetation sub-model (m2501), 614 sample points were suggested, based on the reduction procedures outlined above. Comparison of relative proportions of land area occupied by the various Strata in m2501 to the percentage of sample points in this sub-model suggests a close agreement (Figure 7). This suggests that along with some manually refined selection, proportionate sampling of the landscape should be possible using the procedure outlined herein.

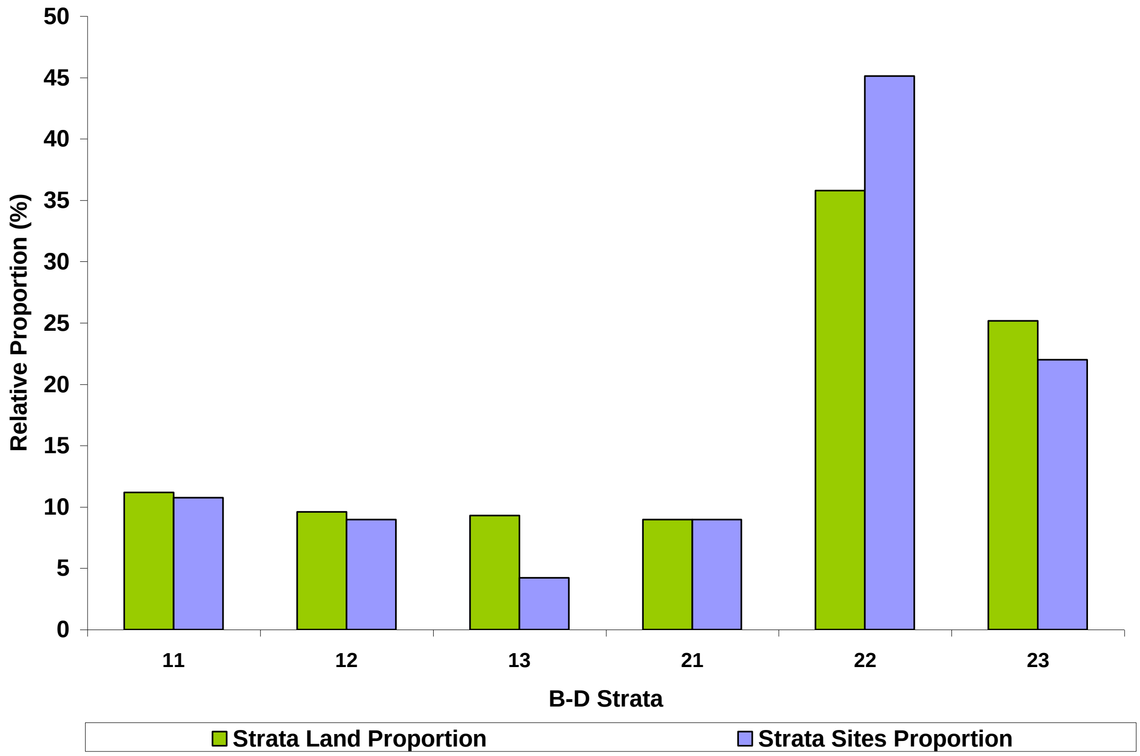


Figure 7. *Proof of concept for proportionate sampling in the B-D NF, using the Tobacco Root Mountains vegetation modeling sub-unit (m2501)*

A further component of the proportionate sampling process that was not included in this analysis is land ownership and access limitations that might be encountered. In concept, access to sample sites on private land may be limited. Taking this into account, sample sites located on private land may be omitted during field activities. In an effort to maintain relative proportions of sites within each of the strata, eliminated sites can be redistributed to more accessible locations within strata of interest.

4. Image Classification

Labeling Algorithms

The Federal Geographic Data Committee (FGDC) Vegetation Classification Standards (FGDC 1997) establishes a hierarchy of existing vegetation classification with nine levels. The top seven levels are primarily based on physiognomy. The two lowest levels, alliance and association, are based on floristic attributes. The USDA Forest Service recently released the national direction for classification and mapping of existing vegetation to implement the FGDC standards and to provide direction for classifying and mapping structural characteristics (Brohman and Bryant 2005). This direction applies to a variety of geographic extents and thematic resolutions characterized as map levels. The Northern Region Vegetation Mapping Project is specifically designed to meet this national program direction at the mid-level.

Most attribute labeling of the VMap products were accomplished using eCognition software. eCognition operates off of a hierarchy classification scheme and within that scheme, a series of functions can be used. For features easily discernable from image statistics (i.e. tree, non-tree; low tree canopy cover, high tree canopy cover), membership functions were used in the hierarchy process to separate cover types. Figure (8) shows an example of one these functions. For features less discernable (dominance type, tree size), membership functions were incorporated with nearest neighbor classification algorithms to provide labeling to the image objects.

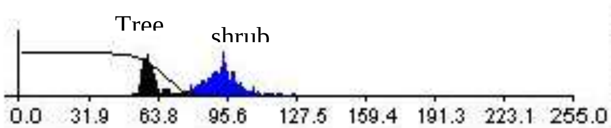
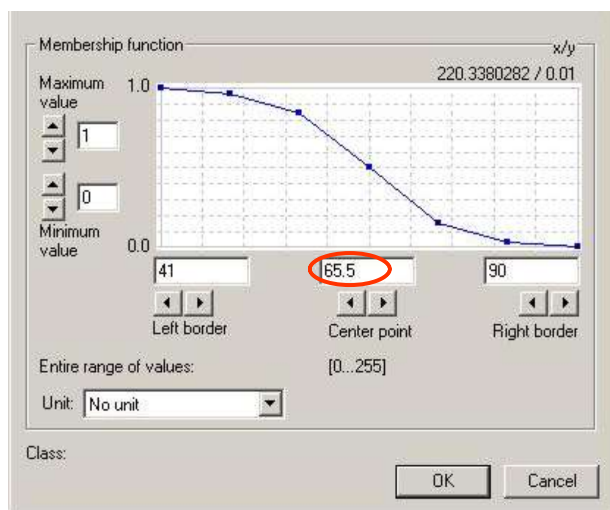


Figure 8. An example of an eCognition member ship function. The example shows 'tree' sample data (blue histogram) and ' shrub' sample data (black histogram) for one of the image inputs. Since there is such good separation between 'tree' and 'shrub' histograms in this example, a membership function is created to separate 'tree' from 'shrub'. At left, the membership excludes 'tree' at 65.5 for this input. A series of these can be created for all image inputs that show separation and combined to create outputs.



The eCognition hierarchy classification scheme sub-divides from general to specific at each level of the hierarchy with the classes at each level inheriting class descriptions from parent classes. The scheme below shows an example of this structure:

LEVEL 1:
- FOREST
- NONFOREST

LEVEL2:
- FOREST
  - 10-40% CANOPY COVER
  - 40%+ CANOPY COVER
- NONFOREST
  - VEGETATED (10% VEGETATED COVER)
  - NON-VEGETATED (LESS THAN 10% VEGETATED COVER)

LEVEL3:
- 10-40% CANOPY COVER
  - LOW CANOPY COVER TREE (10-25%)
  - MODERATE-LOW CANOPY COVER TREE (25-40%)
- 40%+ CANOPY COVER
  - MODERATE-HIGH CANOPY COVER TREE (40-60%)
  - HIGH CANOPY COVER TREE (60%+)
- VEGETATED
  - HERBACEOUS
  - SHRUB
- SPARSELY VEGETATED
  - WATER
  - ROCK

Implementation of this classification hierarchy produces associated geospatial databases for four primary attributes: lifeform, dominance type, tree canopy cover, and tree size class. These original image objects were merged to a 5 acre minimum to produce mid level map products.

Non-forest Labeling Algorithms

The data collected for non-forest cover was done during the field seasons of 2009 and 2010. Although a lot of data was collected overall, there was not sufficient data to drive a separate classification for each individual sub-model using eCognition. To circumvent the problem of not having enough data, the field data was combined for all of the models and classified with 'Random Forests' a classification and regression tree model that is part of the statistical software R version 2.72 program. The algorithm for inducing a random forest was developed by Leo Breiman and Adele Cutler, and "Random Forests" is their trademark (Breiman and Cutler, 2008.) The term came from **random decision forests** that was first proposed by Tin Kam Ho of Bell Labs in 1995. The method combines Breiman's "bagging" idea (Breiman, 1996) and Ho's "random subspace method" (Ho, 1998) to construct a collection of decision trees with controlled variations.

Non-forest classes were taken as far as possible using eCognition. The classes produced for all models in eCognition include: xeric shrub, mesic shrub, dry grass, wet grass, sparse vegetation, and water (see appendix A for a description of these classes.) Additional classes produced from the Random Forests classifier include: bunch grass, single stem grass, two litter classes, and two classes of xeric shrub canopy cover (10-25% and $\geq$ 25%).

Map Product Review

As part of the review process, all models were visited in the field the summer of 2010 and revised based on data collected from that work. This review included only tree attributes since the expanded non-forest classes had not yet been mapped (non-forest data was collected during the review process however.) The field review process is critical for correction of errors associated with the classification and enables a refinement in the final output product that otherwise would not be possible. The resulting classification accuracy numbers (see Section 7) directly reflect the improvement that is seen when adequate field time is allowed. The B-D VMap is the first project completed where two full seasons of field data collection were actually accomplished (on the Eastside VMap almost a full field season was lost to a bad fire year) and the improvement in classification accuracy is dramatic, ranging from a 5% to 23% increase in the four primary classes (DOM60, DOM40, CCV, and TSZ).

Literature Cited

Bailey, R.G., P.E. Avers, T. King, and W.H. McNab, eds. 1994. Ecoregions and subregions of the United States (map). Washington, DC: U.S. Geological Survey. Scale 1:7,500,000; colored. Accompanied by a supplementary table of map unit descriptions compiled and edited by McNab, W. H. and Bailey, R. G. Prepared for the U.S. Department of Agriculture, Forest Service.

Berglund, D., Bush, R., Barber, J., and Manning, M. 2009. R1 Multi-level classification, mapping, inventory, and analysis system. Numbered Report 09-01 v2.0, Missoula, MT: U. S. Department of Agriculture, Forest Service, Region 1.

Breiman L, Cutler A (2008). "Random Forests – Classification Manual (website accessed in 1/2008)." http://www.math.usu.edu/~adele/forests/.

Breiman, L. (1996). Bagging predictors. *Machine Learning 26*(2), 123–140.

Brohman, R. and L. Bryant. 2005. Existing vegetation classification and mapping technical guide. U.S. Department of Agriculture, Forest Service, Washington Office, Ecosystem Management Coordination Staff. http://www.fs.fed.us/emc/rig/documents/integrated_inventory/FS_ExistingVEG_classif_mapping_TG_05.pdf

Coburn C. A. and Roberts A. C. B. A multiscale texture analysis procedure for improved forest stand classification. *International Journal of Remote Sensing, October, 2004*: 25(20), 4287-4308

Federal Geographic Data Committee. 1977. Vegetation Subcommittee. Vegetation classification standard. FGDC-STD-005. Federal Geographic Data Committee, U.S. Geological Survey, Reston, Virginia, USA. [Available online: http://www.fgdc.gov/standards/documents/standards/vegetation/vegclass.pdf]

Federal Geographic Data Committee. 1996. FGDC Standards Reference Model. Federal Geographic Data Committee, U.S. Geological Survey, Reston, Virginia, USA. [Available online: http://www.fgdc.gov/standards/refmod97.pdf]

Gopal, S. and C. Woodcock. 1994. Theory and methods for accuracy assessments of thematic maps using fuzzy sets. *Photogrammetric Engineering and Remote Sensing*, 60: 181-188.

Haralick, R.M. and L.G. Shapiro. 1985. Image segmentation techniques. Comput. *Vis. Image Understand*, 29: 100-132.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, *20*(8), 832–844.

Kauth, R.J. and G.S. Thomas. 1976. The tasseled cap—a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue, University, West Lafayette, IN, 4b: 41-51.

Moore I.D., Gessler P.E., Nielson G.A. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 1993;57:443-452

Press, W. H., S. A. Teukosy, W. T. Vetterling, and B. P. Flannery. 1992. Numerical recipes in fortran: the art of scientific computing, 2nd ed. London: Cambridge, 963 p.

Roberts, D.W., Cooper, S.V. 1989. Concepts and techniques of vegetation mapping. In: Ferguson, D.; Morgan, P.; Johnson, F.D., eds. Land classifications based on vegetation: applications for resource management. Gen. Tech. Rep. INT-257. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station: 90-96.

Ryherd, S. and C. Woodcock. 1996. Combining spectral and texture data in the segmentation of remotely sensed images. *Photogrammetric Engineering and Remote Sensing*, 62: 181-194.

Shandley, J., J. Franklin, and T. White. 1996. Testing the Woodcock-Harward image segmentation algorithm in an area of Southern California chaparral and woodland vegetation. *International Journal of Remote Sensing*, 17: 983-1004.